nature portfolio

Corresponding author(s): Lih Feng Cheow

Last updated by author(s): Jun 5, 2023

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection

Flow-cytometry data were acquired in CytoFLEX (Beckman Coulter) or Moflo Astrios cell sorter (Beckman Coulter).

In high-throughput sequencing studies, the libraries for cDNA, sample hashtags, surface-protein markers and CFSE capture were pooled and sequenced in the Illumina HiSeq X platform.

Data analysis

Flow cytometry data were analysed using FlowJo software v10.5.

The R environment for statistical computing and graphics is version 4.2.2.

Histogram and bar plots with statistical analysis were performed by using GraphPad Prism software v8.0.2.

Raw cDNA FASTQ data were processed for cell-barcodes calling and UMI counting of the cDNA library using Cell Ranger v4.0 (10x Genomics) with default parameters. The generated cell barcodes were used to find individual cell-barcode-matched expression profiles of sample hashtags, surface proteins and cytokines using CITE-seq-Count (v1.4.3) with a Hamming distance of 1 for both cell-barcodes calling and UMI counting. Pre-processed count matrices were imported into the Seurat package (v4.3.0). Dead cells were removed by filtering out cells with > 10% mitochondrial transcript counts, and additional low-quality cells were filtered out by setting gene features > 1,200 but < 8,000, and total gene counts > 1,000 but < 40,000.

mRNA-derived cDNA counts were normalized using log-normalized counts, whereas protein-derived barcodes were normalized using a centered-log-ratio transformation, both of which were embedded in the Seurat package (v4.3.0). Differential gene expression was performed by using the FindAllMarkers() function in Seurat (v4.3.0) with default parameters and min.pct = 0.2. Only genes with adjusted P < 0.05 were

used for downstream analysis. Cell-cycle scoring and regression were performed by functions embedded in the Seurat package (v4.3.0). GSEA was performed using R programs embedded in the clusterProfiler (v3.18.0) through the use of curated hallmark gene sets from the Molecular Signatures Database (MSigDB, v7.4). Where indicated, the capture of CFSE at single-cell level was projected onto the UMAP plot using Single-Cell Signature Explorer (v3.5).

A Naïve Bayes classifier based on module scores of clusters 0 and 1 (dimension reduction by UMAP) was trained on 70% of the cells from clusters 0 and 1 and validated on 30% of the remaining cells. This classifier was applied to cells in cluster 2 to predict their origins.

The code to reproduce all analyses and data in the figures is available at Zenodo (10.5281/zenodo.8004120).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Curated hallmark gene sets are publicly available from Molecular Signatures Database (MSigDB, v7.4). Sequencing data from PAINTKiller-seq is available from the GEO database, with accession number GSE207508. All data needed to evaluate the conclusions described in this paper are included within the paper and its Supplementary Information. The raw and analysed datasets generated during the study are available for research purposes from the corresponding authors on reasonable request. Source data for the figures are provided with this paper.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	No information on sex and gender was available. The residual apheresis blood samples were obtained from anonymous healthy donors who had been given consent for their blood to be used for research at the Health Science Authority (HSA) of Singapore.
Reporting on race, ethnicity, or other socially relevant groupings	No information on race, ethnicity or other socially relevant groupings was available.
Population characteristics	Not available.
Recruitment	The residual apheresis blood samples were obtained from anonymous healthy donors who had been given consent for their blood to be used for research at the Health Science Authority (HSA) of Singapore.
Ethics oversight	This study was approved by the Institutional Review Board of National University of Singapore (NUS-IRB no. H-18-038E).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The sample size was determined by the number of single cells that could be retrieved in FACS and single-cell sequencing. We obtained thousands of cells, which provide sufficiently large sample for our analysis.
Data exclusions	Pre-processed count matrices were imported into the Seurat package (v4.3.0). Dead cells were removed by filtering out cells with > 10% mitochondrial transcript counts, and additional low-quality cells were filtered out by setting gene features > 1,200 but < 8,000, and total gene counts > 1,000 but < 40,000. Samples were demultiplexed, and only singlets with confident specific hashtag staining were included for downstream analysis.
Replication	For the PAINTKiller assay and PAINTKiller-sort, experiments were performed at least twice, independently. Results with at least three

Replication	replicates are shown in the paper, with error bars as indicated. All attempts at replication were successful. For PAINTKiller-seq, a total of two lanes from the pooled DNA libraries were sequenced.
Randomization	No randomization and control of covariates were performed because there was no subjective allocation of samples into experimental groups. Single-cell analysis procedures process data and identify cell algorithmically. For the cell-culture experiments, different conditions were compared and performed, and the data were processed in the same way.
Blinding	Blinding design was not applicable to the study, because the results and data processing were based on objective and quantitative methods (such as flow cytometry and high-throughput sequencing).

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	All primary antibodies (except those home-made conjugates as described below) used in this study are commercially available and have been listed (with supplier name and catalog number) in Methods. These include purified anti-CD45 monoclonal antibody (BioLegend, 304002), polyclonal anti-FITC-biotin (Southern Biotech, 6400-08), CD56-BV786 (BD Biosciences, 564058), CD107a-AF647 (Biolegend, 328612), IFN-γ-PE-Vio770 (Miltenyi Biotec, 130-113-494),anti-FITC-PE (Southern Biotech, 6400-09), as well as a list of DNA-barcoded antibodies from BioLegend for PAINTKiller-seq, including β2M-targeting hashtags (Biolegend TotalSeq A0251 , A0252, A0253), PE (408109, clone PE001), CD3 (300475, clone UCHT1), CD56 (362557, clone 5.1H11), CD16 (302061, clone 3G8), CD57 (393319, clone QA17A04), CCR7 (353247, clone G043H7), CD45RA (304157, clone HI100), CD107a (328647, clone H4A3), CD69 (310947, clone FN50), NKG2D (320835, clone 1D11), NKp44 (325117, clone P44-8), and PD-1 (329955, clone EH12.2H7). The concentrations of detection antibodies were used according to manufacturer's recommendation except if otherwise stated. This information is provided in Methods. To prepare streptavidin-conjugated bispecific antibodies, unconjugated anti-CD45 monoclonal antibody (BioLegend, 304002; Clone HI30) was linked to streptavidin following the instruction of Lightning-Link Streptavidin Antibody Labeling Kit (Novus Biologicals, 708-0030). Sheep anti-FITC-biotin polyclonal antibody (Southern Biotech, 6400-08) was used for the capture of CFSE.
Validation	All reported antibodies as described above (except those home-made conjugates as described below) are commercially available, and validation materials on the species and applications can be found on the vendor websites. The anti-CD45-Streptavidin conjugate (aCD45-Strep) was validated by its ability to bind cell-surface CD45 and to capture biotin-Atto 565 (Sigma-Aldrich, 92637) before used for effector-cell modification.

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	NK-92MI (CRL-2408) and K562 (CCL-243) cell lines were obtained from the American Type Culture Collection (ATCC).
Authentication	Cell lines purchased from ATCC were authenticated by the vendor. The cell lines were further identified by immunostaining using CD56-BV786 and CD3-BUV395, and detected by flow cytometry.
Mycoplasma contamination	Cell lines purchased from ATCC were tested by the vendor and were absent of mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	For all staining protocols, culture cells were washed with cold PBS (0.5% FBS) before subsequent staining at 4 degree in PBS (2% FBS). Human TruStain FcX reagent (Biolegend, 422302) was used for the blockade of FcR-involved non-specific staining. For absolute cell counting, before flow cytometry data acquiring, 20 µl Precision Count Beads (Biolegend, 424902) were spiked into 100 ul cell suspension.
Instrument	Flow cytometry data was collected using CytoFLEX analyzer (Beckman Coulter) or cell sorting was performed using Moflo Astrios sorter (Beckman Coulter).
Software	Analysis of flow cytometry data was performed using FlowJo software version 10.5.
Cell population abundance	NK cells after co-culture with K562 cells were stained with CD56-BV786 and sorted into CD56high CFSE+, CD56high CFSE– and CD56low CFSE– NK sub-populations as gated in Fig. 3b. Purity of sorted cells (day 0) were not detected post-sorting due to limited cell numbers. The homogeneous cluster of CD56low CFSE– NK-derived cells at day 3 post-sort suggests high purity of the sorted cells (Extended Data Fig. 2). In addition, a total of three independent experiments were performed at different times, and a consistent result from functional testing of sorted NK cell subpopulations was observed, which further indicates the quality of the sorting efficiency (Fig. 3c,d).
Gating strategy	Cells were distinguished from debris and count beads where applicable by FSC-A/SSC-A. Singlets were selected by FSC-A/FSC-H. The detailed gating strategies to identify mixed cell populations are provided.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.